SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

(Established under section 3 of the UGC Act 1956)

Re - accredited by NAAC with 'A' Grade

Founder: Prof.Dr. S. B. Mujumdar, M.Sc.,Ph.D. (Awarded Padma Bhushan and Padma Shri by President of India)

Name: Achyut Shukla
Batch: CS – ML – 1
PRN: 20070122005

Lab Assignment – 1

Sessions – 1

Aim: Introduction to open-source software in ML & ML libraries.

Theory:

1. List down Machine Learning software's. Explain the software's in detail. Give examples of one case study in the software.

Ans: The list of ML software are:

TensorFlow; TensorFlow is an open-source software library developed by Google. It provides a comprehensive ecosystem of tools, libraries, and community resources for building and deploying machine learning models.

PyTorch; PyTorch is an open-source deep learning framework developed by Facebook's AI Research lab. It offers a dynamic computational graph, making it easy to build and train neural networks.

Scikit-learn; Scikit-learn is a popular machine learning library in Python, built on top of NumPy and SciPy. It provides a wide range of algorithms and tools for various ML tasks, such as classification, regression, clustering, and dimensionality reduction.

Microsoft Azure Machine Learning: Microsoft Azure Machine Learning is a cloud-based platform that provides a wide range of tools and services for machine learning development and deployment. It offers a drag-and-drop interface for building ML pipelines, supports multiple programming languages, and integrates with other Azure services.

Case Study

The use of Azure Machine Learning for predictive maintenance in manufacturing is a notable case study. By applying ML algorithms to sensor data collected from machines on the

production line, models can be trained to detect anomalies and predict equipment failures in advance. This enables proactive maintenance, reducing downtime and improving overall productivity.

2. List down Machine Learning Libraries. Explain the libraries in 2-3 sentences. Give examples of the libraries

NumPy: NumPy is a fundamental library for numerical computing in Python. It provides a powerful N-dimensional array object, along with a collection of mathematical functions and tools for manipulating arrays. NumPy is widely used in machine learning for data preprocessing, mathematical operations, and array-based computations.

Pandas: Pandas is a Python library that offers data manipulation and analysis tools, particularly for structured data. It provides a high-performance, easy-to-use data structure called the Data Frame, which allows for data cleaning, filtering, grouping, and other data preprocessing tasks.

SciPy: SciPy is a library built on top of NumPy and provides additional scientific computing functionality. It includes modules for optimization, linear algebra, signal processing, image processing, statistics, and more.

Kera's: Kera's is a high-level neural networks API written in Python. It provides a user-friendly interface for building and training deep learning models, with support for various neural network architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and more.

Scikit-learn: Scikit-learn is a comprehensive machine learning library in Python. It offers a wide range of algorithms and tools for tasks like classification, regression, clustering, dimensionality reduction, and model evaluation.

Environment:

1. Python (Version compatible to your OS)
2. Google Collab
3. Jupiter Notebook

Task:

1. Install environment you will be working on for the rest of the semester. Check workability of the environment
2. Analyse the dataset – Mall Customer Dataset (Available on Kaggle/GitHub). Segment the customers based on the age, gender, interest.

Submission :

Word Document with all the above-mentioned points. For Analysis output snapshot is necessary. Last page will contain Plagiarism report. Plagiarism must be less than 10%.

Output & Conclusion:

We have performed Hierarchical and K-means clustering between Age and Spending Score and Annual Salary and Spending score to segment the customers database.

Segmentation using Age and Spending Score

S.no	Clustering Model	Silhouette Score	Calinski Harabasz Score
1	K-Means Clustering	0.5	339.9
2	Hierarchial Clustering	0.46	214.11

Segmentation using Age and Spending Score

S.no	Clustering Model	Silhouette Score	Calinski Harabasz Score
1	K-Means Clustering	0.09	58.1
2	Hierarchial Clustering	0.37	86.79